Report #3
PD-Internship
Lerner Paul
21/03/2018

Introduction

On March 7th and 12th we agreed that it would be interesting that I get into papers about :

- automatic diagnosis of other diseases based on HaW features
- automatic diagnosis of PD using other features than HaW, e.g. speech
- automatic diagnosis (via HaW ?) of non-PD disease that cause micrographia, dysgraphia, bradykinesia or tremor and for Atypical Parkinsonian Syndromes.

So this is what I did this week.
Search equations for the first point : ¬"Parkinson" & ("Handwriting" | "Drawing" | "Dysgraphia" | "Micrographia") &

- "Disease Diagnosis"
- "Automatic Diagnosis"
- "Automated Diagnosis"
- "Decision Support"
- "Machine Learning Diagnosis"
- "Classification Diagnosis"

Search equations for the last 2 points (on Google Scholar) : <Disease Name> &

- "Handwriting Decision Support"
- "Machine Learning"
- "Classification"
- "Automated Diagnosis"

With <Disease Name> being :

- "Progressive Supranuclear Palsy"
- "Corticobasal Syndrome"
- "Huntington"
- "Essential Tremor"
- "Alzheimer"
- "Focal Basal Ganglia Lesions"
- "Parkinson"

- "Multiple System Atrophy"

From the first search equations, I was not able to find any works about diseases other than those listed above.

I also started coding (cf. Gitlab commits) a LSTM using Pytorch[1]. In the future weeks I think I should read more about RNN, LSTM and time series (or maybe about handwriting analysis first ?). Data representation / model training raise a lot of questions (Cf. Conclusion of Report #2). New questions about LSTM are :

- what is a *stateful* LSTM ?
- what model should we use to decode the LSTM output ?
    - n° of layers
    - layers size
    - dropout ?
    - attention ? (though I think it's specific to Seq2Seq)
- should the encoder and the decoder have one optimizer each (so they can have different learning rates) as in this tutorial ?
- should we use LSTM or a GRU ? Leo (thésard de Chloé) told me that LSTM handled long inputs better than GRU
- should we represent the target as a scalar (→ binary classification / sigmoid) or as a one-hot vector (→ multi-class classification / softmax). I heard that one hot works better in practice. Moreover, it allows for simple calibration[2].

Summary

[1] https://pytorch.org/
[2] https://geoffpleiss.com/nn_calibration

Parkinson Disease (PD)

Motivation

"The diagnosis of PD is still largely a clinical one, as there is no definitive test able to confirm the diagnosis during life, with the exception of gene testing in a reduced number of cases." (Massano & Bhatia, 2012)

Moetesum et al. explain that *"Traditional diagnostic procedures for determination of the disease include costly, invasive methods like SPECT and CT scans, which are usually effective when the **disease has already progressed to a mature stage**. Clinical practitioners therefore, first opt for manual, non-invasive screening tests like Unified Parkinson's Disease Rating Scale (UPDRS), for **early detection** of the disease. While this process is quite established and has been modified over years of experience, it remains relatively **subjective**."*.

Wu et al. 2011 state that *"The pathological process leading to PD begins decades before the typical motor symptoms and, by the time the diagnosis is made, about **70% to 80%** of striatal dopamine (DA) and at least **one-third** of substantia nigra (SN) neurons and striatal dopaminergic fibers are already lost."* (cf. article for references) which shows how important it is to **diagnose PD early**.

In the description of the disease made by James Parkinson in 1817, writing deficits precede walking deficits : *"Hitherto the patient will have experienced but little inconvenience; and befriended by the strong influence of habitual endurance, would perhaps seldom think of his being the subject of disease, except when reminded of it by the unsteadiness of his hand, whilst writing or employing himself in any nicer kind of manipulation. But as the disease proceeds, similar employments are accomplished with considerable difficulty, the hand failing to answer with exactness to the dictates of the will."*. This motivates the idea of using handwriting for early diagnosis of PD.

Figure by Letaneux et al. which shows how little handwriting was studied in 2014 : only 2% ! (todo update ?)

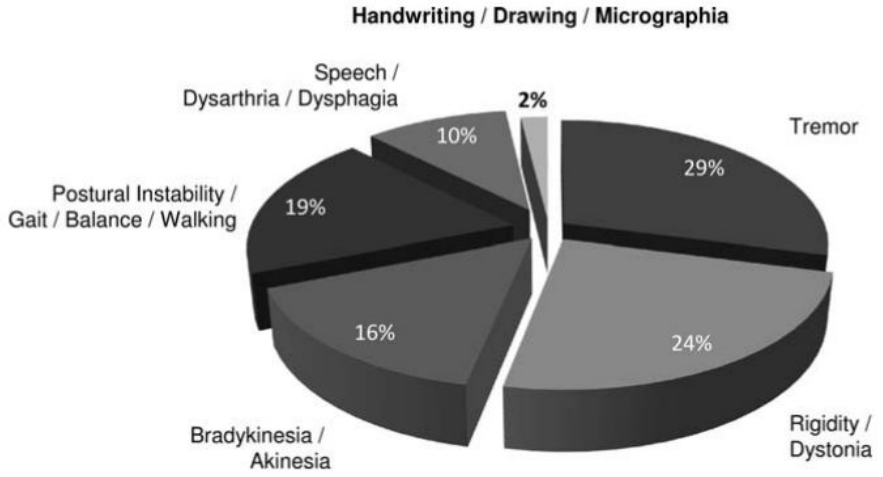

FIG. 1. Search for articles (PUBMED, Feb, 2014) when using the [Parkinson's Disease + key-word #2] combination. Key-word #2 corresponded to the main symptoms or altered functions in PD. The "humans" and "English" filters were used, and for the search involving several keywords #2, the "OR" function was used.

Causes

De Lau & Bredeler state that the causes of the disease are *"still largely unknown"* and that *"older age and smoking habits are the only risk factors for PD that have consistently been found across studies"*.

Biomarkers

Benito-Leon et al. found that patients with Essential Tremor (ET) are 4 times more likely than controls to develop PD (study over 3813 patients).

Wu et al. 2011 wrote a thorough review of PD biomarkers (i.e. early signs of PD). They don't mention micrographia nor dysgraphia, this goes against the findings of McLennan et al. and San Luciano et al. (cf. Report #2). They don't mention dysphonia nor vocal impairment either, which goes against the findings of Harel et al. 2004 who show that vocal disorder may be one of the first symptoms to appear nearly 5 years before clinical diagnosis.

Figure from Wu et al. 2011 :

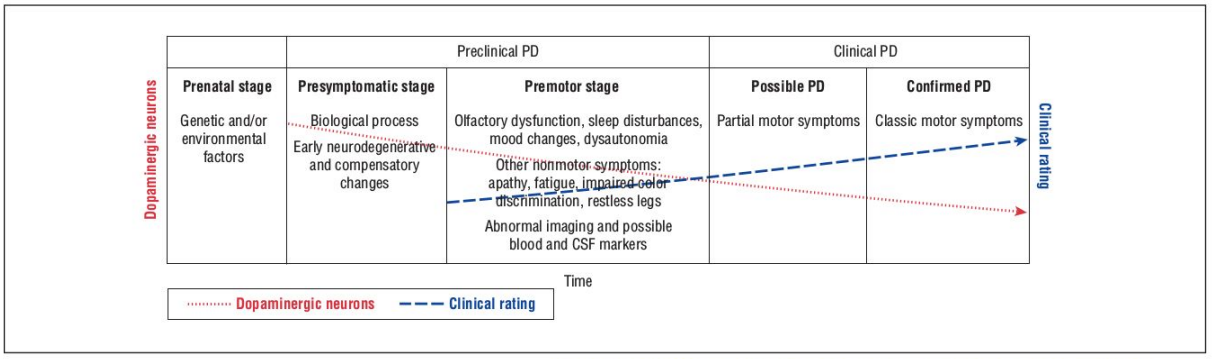

Figure. Hypothesized disease course of preclinical Parkinson disease (PD). The PD hallmark, loss of dopaminergic neurons, develops slowly and gradually progresses over years (preclinical stage). Before cardinal motor symptoms of PD appear, several premotor symptoms are present (premotor stage). The detection of these premotor symptoms may contribute toward an early or preclinical diagnosis of PD. These premotor symptoms may persist in the motor phase. CSF indicates cerebrospinal fluid.

Motor symptoms

"Parkinson's disease is a disease of older age, associated with a loss of dopaminergic cells in the substantia nigra that project to the striatum.This loss causes disturbances of basal ganglia function and results in disturbances of motor control." (Phillips et al. 1991).

Phillips et al. 1991 report that *"Parkinson's disease produces a variety of symptoms; **akinesia, bradykinesia, rigidity, and tremor**. Since these symptoms are relatively **independent** of each other (Zetusky et al. 1985), they probably **involve different mechanisms** (for example, Parkinsonian tremor is associated with cholinergic rather than dopaminergic neurotransmitter systems; Stahl 1986). **The observed handwriting impairments could therefore be the result of a number of symptoms (bradykinesia, rigidity, or tremor).**"*. The authors conclude that, *"given the multi-dimensional nature of Parkinsonian impairments, a **number of specific measures** of handwriting quality are probably necessary to characterize these patients' handwriting."*. This motivates the idea of using deep learning instead of manually extracted features which requires expert intervention (and are therefore subjective / might omit some important ones).

Clinical decision support system (CDSS)

Pereira et al. 2018 wrote a review on "computer-assisted PD diagnosis". It includes works from 2014, 2015 and 2016. They don't make much comparisons between similar works but rather present them independently. Moreover, they omitted several of my references such as San Luciano et al. 2016, Graça et al. (cf. Report #2).

Handwriting analysis

Cf. Report #2 for more details on Clinical decision support system (CDSS) based on HaW analysis. The strongest results (with the largest dataset) are from San Luciano et al. 2016 with 86% Se and 81% Sp. The best results are from Taleb et al. 2017 with 94% Se and 100% Sp.

Speech/Vocal analysis

Chen et al. 2016 wrote a thorough review of CDSS based on speech/vocal analysis. They achieved 98% Se and 91% Sp (10-fold cross validation) using **kernel extreme learning machine** (KELM, Huang et al. 2012) on a 31 subjects (including 23 PD) dataset.

Tsanas et al. 2012 achieved 99% Se and 95% Sp (10-fold cross validation) with a **SVM** on 43 subjects (including 33 PD) dataset using frequency features.

Zuo et al. 2013 achieved 98% Se 97% Sp (10-fold cross validation) using **fuzzy k-nearest neighbor algorithm** on a 31 subjects (including 23 PD) dataset using frequency features.

Exhaled Breath

Tisch et al. 2013 achieved 78% accuracy (cross-validation) between PD and controls on 57 subjects (PD, Alzheimer and controls) using discriminant factor analysis. **No further details as the paper is not free**. They also achieved 84% accuracy **between Alzheimer and PD**.

Electromyography (EMG)

Arvind et al. 2010 used a **RNN** on **power spectral density** (PSD) features extracted from EMG data. In order to ensure stationarity of the signal (to allow for PSD) they segmented the EMG time series is into 1s patterns. Their evaluation is unclear to me. They achieved 96% accuracy. The RNN they used is Elman Network (Marra & Morabito).

Electroencephalography (EEG)

Oh et al. 2018 achieved 85% Se and 92% Sp (stratified 10-fold cross validation) using a **CNN** over a 40 subjects (including 20 PD) dataset. They fed directly the raw data to the CNN without any transformation : they used **1-D convolutions** (as Eskofier et al. 2016) and therefore didn't transform the EEG signals into an image, it'd be interesting to apply the same method to HaW.
They used 20% of the training set to validate the model while applying the cross-validation. I guess we could do that as well (given that they had less data than us) in order to implement early stopping, **if necessary**.

Magnetic Resonance Imaging (MRI)

Rana et al. 2015 achieved 87% Se and 87% Sp (leave-one out cross-validation scheme) on a 60 subjects dataset (including 30 PD) using a **SVM** trained on different features extracted from MRI data. They report that most work usually use MRI in order to make differential diagnosis between PD and other diseases (cf. CDSS on other diseases and Conclusion), therefore, their work is innovative.

Gait analysis

Abdulhay et al. 2018 achieved 97% Se and 87% Sp (unknown method / data split) over 166 subjects (including 93 PD) using an **SVM** on frequency features extracted from vertical ground reaction force. The authors applied **Fast Fourier Transform (FFT)** on the data before extracting the features. It's possible to use FFT for gait analysis because the signal is stationary or subsection stationary, unlike tremor signals (cf. Ai et al. 2011) and HaW signals : *"The digital representation of handwriting as a based time series is the result of several interacting mechanisms like **tremor**, or irregular muscle contractions that introduce **randomness** to the movement during handwriting. [...] In addition, the **nonstationary signals** have statistical properties that vary as a function of time and should be analyzed differently than stationary data"* (Taleb et al. 2017).

CDSS on other diseases

Essential tremor (ET)

Ai et al. 2011 achieved 97.5% Se and 98.33% Sp (4-fold cross validation) classifying PD and ET. They used a **SVM** on features extracted from **empirical mode decomposition (EMD).** The dataset consists of acceleration signals of 25 subjects.

Cf report #2 : Yu et al. reported significant differences in handwriting kinematics between patients with PD and Essential Tremor (ET)

Alzheimer disease (AD)

Tisch et al. 2013 achieved 85% accuracy classifying AD and controls (cf. Exhaled Breath).

Liu et al. 2015 achieved 89% Se and 87% Sp (10-fold cross validation) classifying AD and controls on MRI images from 311 subjects (including 65 AD and 77 controls). They used an auto-encoder in order to learn representation from the data. They then classified AD and controls by connecting a softmax layer to the last hidden layer of the auto-encoder. They also used a SVM in order to compare the results : the auto-encoder based neural network outperformed the SVM (same Sp but 4% better Se).

Huntington's disease (HD)

Perez et al. 2018 worked on utterance data extracted from speech from 62 subjects (including 31 HD). Among the 31 HD, 11 are premanifest, 12 are in the early stage, and 8 are in the late stage. The authors compared four models :

- **k-NN** with Euclidean distance (which they fed speaker-level features)
- k-NN with DTW distance (which they fed utterance-level features)
- Deep Neural Networks (**DNN**, without convolutions, which they fed speaker-level features)
- Long-Short-Term Memory Recurrent Neural Networks (**LSTM**, which they fed utterance-level features).

Their LSTM is comprised of two LSTM layers with recurrent dropout, bias l2 regularization, and kernel l2 regularization followed by a softmax output layer (isn't sigmoid the usual activation function for binary classification ?). The authors achieved similar results among the three models : **87%** accuracy (Leave-One-Subject-Out paradigm). Though we advise the reader that the authors use an ensemble approach, in which they train five separate models and the *"mode"* of the five predictions is used as the final prediction of the system. Approach which a priori always enhances the results **[REF]**. They also present their results depending on the disease's stage in this confusion matrix :

	Healthy	HD
Healthy	0.95	0.05
Premanifest	0.54	0.46
Early	0.14	0.86
Late	0.02	0.98

Their results motivate the idea of using LSTM for our work although the authors didn't fed raw data to the LSTM but utterance-level features (a vector of different features for each utterance of a given speech).

Sarbaz et al. 2014 achieved 96.6% accuracy using an **artificial neural network** trained on features extracted using **power spectral density (PSD)** on gait data. *"This system can diagnose patients at the first stages of the disease, and it also can recommend suspected persons to the specialist."*. **No further details as the paper is not free**.

Klöppel et al. 2009 used a **SVM** to automatically identify presymptomatic HD gene mutation carriers (PSCs) in the absence of any a priori information. They worked on a 191 subjects dataset (including 96 PSC). Using the gray matter segment of MRI scans, they achieved 69% accuracy on subjects with at least a 33% chance of developing unequivocal signs of HD in 5 years.

Multiple System Atrophy (MSA)

Cf. Progressive Supranuclear Palsy (PSP) for Scherfler et al. 2016 work review.

Progressive Supranuclear Palsy (PSP)

Scherfler et al. 2016 provide *"Class III evidence that automated MRI analysis accurately discriminates among early-stage PD, MSA, and PSP"*. Their dataset contains MRI data from 40 PD, 40 MSA, and 30 PSP (totalizing 110 subjects). *"The diagnostic accuracy for PD vs MSA or PSP was **97.4%**. In contrast, diagnostic accuracy based on validated clinical consensus criteria at the time of MRI acquisition was **62.9%**."*. **No further details as the paper is not free**.

Cf Report #2 : Ling et al. show that micrographia is more frequent in Progressive Supranuclear Palsy (PSP) than in PD and that finger tap analysis is a better tool to distinguish the two diseases.

Conclusion

Note : **without extensive research**, it seems that most of CDSS for differential diagnosis (e.g. between PD and PSP) are based on **MRI** data. HaW might not be appropriate for differential diagnosis, this is suggested by Ling et al. work (Cf. Report #2).

Most works on ET, PSP and MSA seems to be differential diagnosis between those and PD.

SVM seems clearly to be the most widely used algorithm to date. RNN doesn't seem to have been widely explored.

About Fourier transform, power spectral density (PSD), empirical mode decomposition (EMD) and Intrinsic Mode Functions (IMF) :

- PSD comes from Fourier transform, therefore if one can't apply Fourier, one can't apply PSD
- Fourier is only applicable on **stationary signals** such as vertical ground reaction force (cf. Gait analysis).
- Thus, Fourier is not applicable on **non-stationary** signals, such as:
    - Handwriting
    - tremor
    - EMG
- although one trick is to segment the data in X second patterns if it is **subsection stationary** (cf. Electromyography (EMG)).
- For **non stationary signals** we can use Hilbert-Huang Transform (HHT) which uses the EMD method to decompose a signal into intrinsic mode functions (IMF) with a trend, and applies the HSA method to the IMFs to obtain instantaneous frequency data. According to Taleb et al. 2017 : *"the first few IMFs contain only time-varying high spectral components representing **the noise** which is believed to reflect the presence of PD. **The noise in the signal seems to corresponds to tremor or jerk** and therefore to identify the presence of the disease."*.

So, to summarize, if I understood correctly, in our case (HaW + RNN), we can only do as Arvind et al. 2010 (cf. Electromyography (EMG)) if we assume that HaW is subsection stationary (but Ai et al. 2011 make me think the opposite).

Todo List

Cf. Introduction.

References

Abdulhay, E., Arunkumar, N., Narasimhan, K., Vellaiappan, E., & Venkatraman, V. (2018). Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease. Future Generation Computer Systems, 83, 366-373.

Ai, L., Wang, J., & Yao, R. (2011). Classification of parkinsonian and essential tremor using empirical mode decomposition and support vector machine. Digital Signal Processing, 21(4), 543-550.

Arvind, R., Karthik, B., Sriraam, N., & Kannan, J. K. (2010, October). Automated detection of PD resting tremor using PSD with recurrent neural network classifier. In *2010 International Conference on Advances in Recent Technologies in Communication and Computing* (pp. 414-417). IEEE.

Chen, H. L., Wang, G., Ma, C., Cai, Z. N., Liu, W. B., & Wang, S. J. (2016). An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease. Neurocomputing, 184, 131-144.

Eskofier, B. M., Lee, S. I., Daneault, J. F., Golabchi, F. N., Ferreira-Carvalho, G., Vergara-Diaz, G., ... & Bonato, P. (2016, August). Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 655-658). IEEE.

Harel, B., Cannizzaro, M., & Snyder, P. J. (2004). Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. *Brain and cognition*, *56*(1), 24-29.

Klöppel, S., Chu, C., ... & PREDICT-HD Investigators of the Huntington Study Group. (2009). Automatic detection of preclinical neurodegeneration: presymptomatic Huntington disease. Neurology, 72(5), 426-431.

De Lau, L. M., & Breteler, M. M. (2006). Epidemiology of Parkinson's disease. *The Lancet Neurology*, *5*(6), 525-535.

Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., & Feng, D. (2014, April). Early diagnosis of Alzheimer's disease with deep learning. In 2014 IEEE 11th international symposium on biomedical imaging (ISBI) (pp. 1015-1018). IEEE.

Massano, J., & Bhatia, K. P. (2012). Clinical approach to Parkinson's disease: features, diagnosis, and principles of management. *Cold Spring Harbor perspectives in medicine*, *2*(6), a008870.

Neils-Strunjas, J., Groves-Wright, K., Mashima, P., & Harnish, S. (2006). Dysgraphia in Alzheimer's disease: a review for clinical and research purposes. *Journal of speech, Language, and Hearing research*, *49*(6), 1313-1330.

Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., & Acharya, U. R. (2018). A deep learning approach for Parkinson's disease diagnosis from EEG signals. Neural Computing and Applications, 1-7.

Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *42*(2), 513-529.

Marra, S., & Morabito, F. C. (2006). A new technique for solar activity forecasting using recurrent Elman networks. *International Journal of Computational Intelligence*, *3*(1), 8-13.

Parkinson, J. (1817). An essay on the shaking palsy: London: Whittingham and Rowland for Sherwood. *Neely and Jones*.

Pereira, C. R., Pereira, D. R., Weber, S. A., Hook, C., de Albuquerque, V. H. C., & Papa, J. P. (2018). A survey on computer-assisted Parkinson's Disease diagnosis. Artificial intelligence in medicine.

Perez, M., Jin, W., Le, D., Carlozzi, N., Dayalu, P., Roberts, A., & Provost, E. M. (2018, January). Classification of Huntington Disease using Acoustic and Lexical Features. In Proceedings of INTERSPEECH (Vol. 2018, pp. 1898-1902).

Phillips, J. G., Stelmach, G. E., & Teasdale, N. (1991). What can indices of handwriting quality tell us about Parkinsonian handwriting?. *Human Movement Science*, *10*(2-3), 301-314.

Rana, B., Juneja, A., Saxena, M., Gudwani, S., Kumaran, S. S., Agrawal, R. K., & Behari, M. (2015). Regions-of-interest based automated diagnosis of Parkinson's disease using T1-weighted MRI. Expert Systems with Applications, 42(9), 4506-4516.

Sarbaz, Y., & Pourhedayat, A. (2014). Spectral analysis of gait disorders in Huntington's disease: a new horizon to early diagnosis. *Journal of Mechanics in Medicine and Biology*, *14*(01), 1450001.

Scherfler, C., Göbel, G., Müller, C., Nocker, M., Wenning, G. K., Schocke, M., ... & Seppi, K. (2016). Diagnostic potential of automated subcortical volume segmentation in atypical parkinsonism. *Neurology*, *86*(13), 1242-1249.

Tisch, U., Schlesinger, I., Ionescu, R., Nassar, M., Axelrod, N., Robertman, D., ... & Haick, H. (2013). Detection of Alzheimer's and Parkinson's disease from exhaled breath using nanomaterial-based sensors. *Nanomedicine*, *8*(1), 43-56.

Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Transactions on biomedical engineering*, *59*(5), 1264-1271.

Wu, Y., Le, W., & Jankovic, J. (2011). Preclinical biomarkers of Parkinson disease. *Archives of neurology*, *68*(1), 22-30.

Zuo, W. L., Wang, Z. Y., Liu, T., & Chen, H. L. (2013). Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach. *Biomedical Signal Processing and Control*, *8*(4), 364-373.